Use of molecular markers to contribute to the identification of priority coconut accessions in the framework of the Global Strategy for Conservation and Use of Coconut Genetic Resources.

(LoA no. 2013/25)

Final report

Luc Baudouin

BIOVERSITY INTERNATIONAL REPORT FACT SHEET

Library Report No. (Library use only)	

FULL TITLE of REPORT/PROJECT	"Use of molecular markers to contribute to the identification of priority coconut accessions in the framework of the Global Strategy for Conservation and Use of Coconut Genetic Resources."	
AUTHOR of *Commissioned Organization or Contractor (Name and title of person)*	Luc Baudouin, geneticist	
NAME/ADDRESS of *Commissioned Organization or Contractor*	CIRAD - AGAP Avenue Agropolis 34398 Montpellier Cedex 5	
DATE REPORT SUBMITTED	05/02/2014	
TYPE OF REPORT i.e. progress/final	final	
PREVIOUS REPORTS **(Please fill in dates and add more lines if necessary)** **To be completed by BIOVERSITY INTERNATIONAL Coordinator**	1st Progress Report	Date:01/11/2013
	2nd Progress Report	Date:
	3rd Progress Report	Date:
	Final Report	Date:
BIOVERSITY INTERNATIONAL LETTER OF AGREEMENT To be completed by BIOVERSITY INTERNATIONAL Coordinator	LOA 2013/25	
BIOVERSITY INTERNATIONAL PROJECT CODE **To be completed by BIOVERSITY INTERNATIONAL Coordinator**	CRP 6	
BIOVERSITY INTERNATIONAL CONTACT	Dr Roland Bourdeix	

ABSTRACT *(Minimum 100 words)*	To assist germplasm management using molecular markers, we propose here a method to be used in conjunction with geographic and phenotypic information. It can be used to evaluate the number of accessions that need to be reproduced as safety duplication or triplication. Current knowledge of coconut diversity suggests that 60 Tall or "compact" accessions are sufficient. The additional surface needed to accommodate this safety collection is about 60 ha. Most homozygous Dwarf cultivars could be preserved in small numbers. A complementary microsatellite analyses should involve - cultivars/populations already analyzed with insufficient numbers (318 individuals), - analyses in "unchartered areas" of the coconut genetic diversity. map (860 individuals in 10 years), - country-wide studies in Sri Lanka and in the Philippines (with possible extension to India and Indonesia – 200 individuals per country) . To make microsatellite data more available, we improved the user interface of TropGENE-db, introduced more data (554 individuals) and invited 4 countries to input data from about 500 individuals.. Finally, we recommend that, once more powerful genotyping methods become available, a representative subset of 50 to 100 individuals analyzed with into analysis programmes.
KEYWORDS	Country/Region: Global Crop(s): Coconut Subject: DNA analysis, genomics, genetic diversity

Contents

Terms of reference

Objective 1: Assist COGENT in developing a long term ex situ conservation policy of coconut germplasm. This assistance will consist in using molecular markers as a way of reducing redundancy in the collection while ensuring that genetic diversity is appropriately represented. The aim is to identify a list of accession to be preserved and studied in priority irrespective of where they are presently conserved, in order to form a "virtual" or "networked" global collection representing coconut genetic diversity in a cost-effective way. (planned for 01/10/2013)

Objective 2: Identify gaps in coconut molecular characterization. The resulting list will include both cultivars/populations[1] that were analyzed with an insufficient number of individuals and geographic locations of interest in the history of coconut evolution and where coconut diversity has not yet been characterized. (planned for 01/10/2013)

Objective 3: Interact and communicate with the International Thematic Action Group on Coconut Genomics and obtain a consensus about points 1 and 2, in order to include the conclusions of these studies in the Global Strategy for Conservation and Use of Coconut Genetic Resources to be published in December 2013. (planned for 01/11.2013)

Objective 4: Assist the Philippines and Sri Lanka in developing their long term ex situ conservation policy of local germplasm. This assistance will consist in using molecular markers as a way of reducing redundancy in the collection while ensuring that genetic diversity at the country level is appropriately represented. (planned for 15/12/2013)

Objective 5: Invite India, the Philippines, Mexico, Brazil, China and Sri Lanka to contribute data to TropgenDB. As regards genotyping data, those obtained with the coconut kit developed by CIRAD should be considered in priority but other data is also welcome. Markers, linkage maps and QTL data are equally welcome. (planned for 15/12/2013)

Objective 6: Make the user-interface of the coconut section of TropgenDB more accessible and complete for international users. This multi-crop TropgenDB WEB database hosted by CIRAD holds molecular data of 1321 coconut individuals and is accessible online to the whole coconut research community. The objective is to add more data and standardized passport information (the conservation site code and unique access number, international cultivar name and population number) for a better identification of the individuals. (planned for 15/12/2013)

---

[1] In the received nomenclature of coconut genetic diversity, a "population" name can be specified after the cultivar name. Most of the times, the population name merely specifies a geographic origin without implying genetic differences. In a few cases, it may imply minor genetic differences which are judged insufficient to justify creating a new cultivar. According to the context, we may use this word with a different meaning eg. "a mendelian population" or "an in situ population" as opposed to a "collection accession".

Work Plan and Deliverables

Dates	Deliverables	Verifiable indicators
01/10/2013	First draft lists of "priority accessions" and "gaps in molecular characterization"	Lists sent to COGENT secretariat and COGENT ITAG Genomics for further discussion and approval
01/10/2013	First draft version of the Annex 7 of the Global Strategy for coconut conservation and use: "Description of the agreed methodology to select accessions based on allelic diversity"	Draft version of Annex 7 sent to COGENT secretariat and to COGENT ITAG Genomics for further discussion and approval
01/11/2013	Final lists of "priority accessions" and "gaps in molecular characterization"	Consensus lists sent to COGENT secretariat and COGENT ITAG genomics.
01/11/2013	Final version of the *Annex 7* of the GSCCU	Final consensus version sent to COGENT secretariat and to COGENT ITAG genomics.
15/12/2013	Updated Coconut data in TROPGEN	Increased coconut data set with additional passport data (Accession number in the meaning of CGRD database, international name and abbreviation of coconut cultivars)
20/12/2013	Final report	Final report summarizing activities and detailing the data improvement in TROPGEN

Executive summary and recommendations

- We propose here a semi-automatic method to assist germplasm management using molecular markers. This method is to be used in conjunction with geographic and phenotypic information on the concerned populations.
- The underlying principles of this method could also be used to evaluate the number of accessions that need to be reproduced in two or three different gene banks for the purpose of preventing loss of genetic diversity at the species level.
- Given that the number of cross-pollinating cultivars strictly necessary to cover all the diversity of the species, we suggest that 60 Tall or "compact" accessions represented at three different locations are sufficient. The additional surface needed to accommodate this safety collection is about 60 ha.
- Regarding Dwarf, we recommend to conserve every Dwarf cultivar provided it is recognized as different. But small samples (5 to 10 individual /accession).
- The above recommendations correspond to what is specifically needed for the safety preservation purpose at an international level. Breeding and seed production activities are not included.
- We propose to complement microsatellite analyses made on cultivars/populations already analyzed with insufficient numbers. It would represent some 318 individuals from 40 collection accessions and 2 *in situ* populations. This would provide a more reliable estimation of genetic diversity parameters.
- In addition, we propose to reduce unchartered areas in the map of coconut genetic diversity. The whole program would require analyzing 860 individuals (including 100 individuals representing 10 collection accessions. Given that most of the sample collections will be made outside of genebanks (and, sometimes, in remote places) this programme is somewhat demanding and should be planned for the next 10 years.
- We propose to conduct two studies at the country level, involving a total of 400 individuals in Sri Lanka and in the Philippines. (if judged useful, this programme could be doubled by including India and Indonesia.
- We invited 4 countries to input genotyping and marker data into TropGENE-db, a freely accessible database maintained by CIRAD. Potentially, this would represent more than 500 more individuals genotyped with the COGENT microsatellite kit.
- In parallel, CIRAD improved the interface of TropGENE-db in order to make it more informative and easier to use. We entered genotypic data of 554 individuals analyzed with 13 markers.
- Finally we acknowledge Prof. Heslop Harrison's suggestion (cf appendix 6) that the microsatellite kit will not last forever. While we consider that the programmes proposed in the present document remain their value, we recommend that, once more powerful methods are available, a representative subset of 50 to 100 individuals already analyzed with SSRs, should be included into the analyses conducted with the new methods.

Introduction

The present report comes in line with previous work conducted by CIRAD in the framework of the IPGRI-COGENT project LOA 99/079 (the development of a microsatellite kit for use with coconuts.) and the Generation Challenge programme tier 2. The first of these proojects led to An initial set of 14 microsatellite markers was developed and tested with 571 individuals. The size of the data matrix was extended to 1215 individuals times 30 markers during the tier2 programme. Finally we proposed an evolution of the kit and complementary analyses in the report "Consolidate microsatellite data on coconut diversity" elaborated in consultancy conducted under LOA APO/08/007. The present report was requested by the COGENT coordinator to provide elements to the "Global Strategy for the Conservation and Use of Coconut Genetic Resources"

An interim report was submitted on 01/10/2013. This is the final report.

Activities conducted

Objective 1

"Assist COGENT in developing a long term ex situ conservation policy of coconut germplasm. This assistance will consist in using molecular markers as a way of reducing redundancy in the collection while ensuring that genetic diversity is appropriately represented. The aim is to identify a list of accession to be preserved and studied in priority irrespective of where they are presently conserved, in order to form a 'virtual' or 'networked' global collection representing coconut genetic diversity in a cost-effective way".

Allele diversity studies as a methodology to assist accession selection.

Disclaimer: This section is intended to be the basis of annex 7 of the coconut global strategy document whose initial title was: *Description of the agreed methodology to select accessions based on allelic diversity.* It must be noted however that decision about conservation, evaluation and use of genetic resources needs to be made based on a number of factors of which agronomic and technological value should come first. Molecular marker studies may be used to complement this primary information they can by no mean replace them. Thus, the title *Allele diversity studies as a methodology to assist accession selection* would be preferable.

In the same way as molecular pedigree assessment can be used to check the legitimacy of artificial pollination, software *Geneclass* 2[2,3] has been used successfully to check the overall conformity of a reproduced accession. We propose here a method derived from *Geneclass* which aims at reducing unnecessary redundancy and thus at reducing the cost of conservation and of observation of genetic resources. It is intended for cross-pollinating populations (Tall and "compact" Dwarfs in coconut). For the sake of generality, we will refer hereunder to "operational taxonomic units" (OTUs), which, according to the framework of the study, they may represent cultivars in genebank or local populations in a large country and possibly something different in another species). Given that a number of OTUs were observed and characterized, both based on phenotypic and molecular basis. Seeking cost-effective germplasm conservation means that, ideally, we should *represent* each OTU under consideration by *one and only* one member of the reduced set. From another point of view, it can

---

[2] Piry, S., et al., GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of heredity, 2004. 95(6): p. 536-539.

[3] Baudouin, L., S. Piry, and J.M. Cornuet, Analytical bayesian approach for assigning individuals to populations. Journal of heredity, 2004. 95(3): p. 217-224.

be admitted that some degree of redundancy is desirable. It can be incorporated in the present framework by adjusting the number of OTUs retained. This will be discussed in conclusion.

We say that the OTU A is said to represent adequately the OTU B if its probability of producing genotypes similar to those observed in the OTU B is large enough. The method we propose here consists in adding successively OTUs into the reduced set until all OTUs are represented to a certain level. (or alternatively until a fixed number of OTUs is reached).

The typical scenario we consider is the following: a number of OTUs have been observed in a research station (or surveyed in the field). Based on these observations, three sets can be formed:

A: Chosen. OTUs that need to be in the in the reduced set because they are especially attractive or because they represent a standard in genetic trials.
B: Excluded. OTUs that are not available to form the reduced set but that are still to be represented.
C: Candidates. All the other OTUs.

The next steps consist in moving OTUs from set C into set A, based on the assignment score provided by Geneclass 2. In fact, this score is precisely $-\log P_{ij}$ where P_{ij} is the probability of obtaining the genotypes of OTU j (from set C), given the allele frequencies in OTU i (from set A). The "representation score" S_j of OTU j is defined as $\min(-\log P_{ij})$ with i describing all OTUs of set A. The lower the score is, the better the representation. The proposed rule is the following: at each step, move the OTU which is the most poorly represented by set A i.e. with the highest S_j. We wrote an R script to apply this rule[4]. The procedure is described more in detail in appendix 5.

The proposed procedure is semi-automatic: at each step the user has a choice to move the OTU proposed by the R procedure, to move another one or to stop, based on the information provided. In fact, three stopping criteria can be used:

the size of the A subset reached a predetermined value (fixed by economic considerations),
each OTU of the initial set is adequately represented,
the procedure tends to propose OTUs that are known to be very similar to OTUs already in the A subset.

Finally, the procedure provides quality indicators

reduction of the size of the collection
quality of representation (as indicated by low values of S_j)
reduction of redundancy (as indicated by high values of $-\log P_{ij}$ among members of subset A
in addition, the best representative (in set A) of a OTU from subsets B and C is indicated.

We present here an example of the application of the proposed procedure to the cultivars genotyped in the GCP study for illustrative purposes only. Grey cells indicate that a population of the reduced set (columns) is the best representative of a population of the full set (rows). The restricted reference set has only 19 populations and provide an accurate of most populations: the score is between 0 and 5 for 59 populations and

[4] For convenience, the well-known R statistical software was used. If necessary, this script could easily be converted into a stand-alone programme.

doesn't exceeds 11[5]. Redundancy has been drastically reduced: among members of the reduced set (rows in bold), scores generally exceed 10 most of the time.

Table 1: Subset for the Panama, Micronesia and Southeast Asian cultivars

	PNT	TUVT	PADT	TAGT	THT	
PNT	**0.0**	**39.9**	**16.7**	**22.8**	**25.5**	**Panama Tall**
PNTa	3.4	40.2	16.3	23.3	25.1	Panama Tall Aguadulce
PNTcr	10.1	41.4	20.1	29.0	29.7	Panama Tall Costa Rica
KIT	42.7	10.6	19.3	26.9	19.1	Kiribati Tall
TUVT	**48.9**	**0.0**	**23.3**	**31.1**	**17.4**	**Tuvalu Tall**
MIT	40.2	8.9	22.4	29.8	19.0	Marshall Island Tall
BALT	34.6	20.5	7.3	11.5	9.4	Bali Tall
BAYT	30.8	25.0	8.2	13.5	16.4	Bay Bay Tall
MXPT	34.6	22.3	11.0	15.4	11.9	Mexican Pacific Tall
PADT	**27.5**	**22.3**	**0.0**	**8.2**	**6.8**	**Pandan Tall**
SNRT	28.2	21.5	2.6	9.0	3.8	San Ramon Tall
MACT	30.5	27.1	8.6	11.2	12.0	Macapuno Tall
TAGT	**40.3**	**34.6**	**14.0**	**0.0**	**16.9**	**Tagnanan Tall**
KAT	41.9	15.1	9.2	15.1	1.4	Cambodia Tall
MLT	42.8	16.4	11.3	19.1	1.9	Malaysian Tall
DAUT	33.6	17.2	6.1	14.0	1.8	Dau Tall
SKT	36.4	16.5	9.9	12.0	0.0	Sarawak Tall
THT	**37.7**	**14.0**	**8.6**	**15.0**	**0.0**	**Thailand Tall**
LPT	43.0	14.0	9.4	14.6	1.0	Lubuk Pakam Tall
TAAT	39.0	21.2	13.4	19.6	2.4	Ta Tall
HAIT	36.2	24.4	14.1	15.4	9.6	Hainan Tall
SAT	36.0	13.8	6.5	10.2	0.0	Sawarna Tall
PUT	37.1	12.2	5.9	6.8	2.5	Palu Tall
TKT	34.5	11.5	6.5	8.8	2.1	Takome Tall

[5] See MXPT in table 1. This cultivar would be the next one to be introduced into the reduced set.

Table 2: Subset for the Melanesian and Polynesian cultivars

	RIT	VTT	MVT	KKT	NLAD	RTMT	TGT	
RIT	**0.0**	**14.0**	**23.1**	**20.3**	**38.0**	**32.9**	**30.7**	**Rennell Island Tall**
SIT	13.6	0.0	5.5	2.3	30.2	20.9	16.1	Solomon Islands Tall
BBRT	16.9	0.0	5.0	3.7	19.5	18.0	8.7	Baibara Tall
NCT	23.1	3.4	11.5	13.4	20.4	14.5	14.1	New Caledonia Tall
VTT	**21.4**	**0.0**	**17.0**	**12.3**	**27.5**	**20.8**	**19.5**	**Vanuatu Tall**
PLT	25.4	3.1	8.0	3.3	29.0	20.3	17.0	Poligolo Tall
KWT 02	26.6	4.6	9.8	6.0	30.9	21.9	13.9	Kiwai Tall
MBT 03	21.9	1.2	4.0	2.0	25.2	20.0	12.4	Milne Bay Tall
HLT	22.6	8.6	7.6	10.0	26.0	18.6	19.2	Hihisu Tall
OLT 01	27.7	8.0	0.0	6.8	32.7	26.1	21.9	Oro Tall
MVT	**26.5**	**11.4**	**0.0**	**6.0**	**33.4**	**25.4**	**18.6**	**Markham Valley Tall**
VLT	23.2	3.9	6.3	3.6	26.5	21.2	13.2	Vailala Tall
MAT	16.8	1.4	6.2	0.2	25.4	15.5	7.2	Manus Tall
ELT	23.0	3.2	5.9	0.0	30.1	22.6	9.9	East Sepik Tall
KKT	**23.8**	**5.4**	**6.4**	**0.0**	**30.3**	**23.7**	**12.4**	**Kar Kar Tall**
Mad R	28.7	4.5	8.1	0.0	23.4	16.8	13.7	Madang Tall Red
Mad Y	20.5	5.5	4.4	0.0	28.1	21.4	6.3	Madang Tall Yellow
WLT	16.7	1.9	4.4	1.2	28.6	21.4	12.2	West New Britain Tall
GPT	25.4	26.9	10.1	5.1	33.9	24.3	17.0	Gazelle Peninsula Tall
NLT	20.2	5.5	4.8	0.1	23.4	19.1	11.5	Namatanai Tall
TONT	24.4	10.8	23.4	22.2	8.3	10.2	20.5	Tonga Tall
NLAD	**42.5**	**23.8**	**30.5**	**31.8**	**0.0**	**18.3**	**33.6**	**Niu Leka Dwarf**
COKT	23.9	6.2	16.1	12.5	10.6	6.4	13.2	Cook Island Tall
RTMT	31.0	8.3	14.4	15.8	16.1	0.0	24.0	Rotuman Tall
TAT	25.8	10.4	15.9	11.5	11.6	4.1	18.2	Tahitian Tall
TGT	**24.3**	**7.4**	**9.9**	**3.7**	**25.7**	**20.5**	**0.0**	**Tonga Tall**

Table 3: Subset for the Indo-Atlantic cultivars

	ADOT	SNRTsl	MZT	MXAT	WCT	WAT	BRT	
ADOT	**0.0**	**17.7**	**6.8**	**21.3**	**7.8**	**28.6**	**20.1**	**Andaman Ordinary Tall**
SNRTsl	**16.7**	**0.0**	**10.9**	**22.3**	**24.3**	**44.6**	**26.1**	**San Ramon Tall Clovis**
CMT	14.2	11.2	0.0	29.4	26.9	42.0	27.4	Comoro Moheli Tall
MZT	**18.9**	**16.7**	**0.0**	**23.2**	**23.0**	**35.3**	**23.0**	**Mozambique Tall**
EAT	11.6	15.2	0.5	20.5	16.7	25.1	17.6	East African Tall
ECT	9.5	10.8	0.7	13.5	10.8	23.8	12.6	Indian East Coast Tall
LCT	11.5	15.1	0.0	15.8	8.3	17.5	14.7	Laccadive Ordinary Tall
LMT	17.7	17.9	6.5	10.9	9.7	9.8	8.4	Laccadive Micro Tall
SLT	15.5	13.9	9.5	14.3	13.8	21.0	18.5	Sri Lanka Tall
MXAT	**21.1**	**22.2**	**15.4**	**0.0**	**15.8**	**27.7**	**20.1**	**Mexican Atlantic Tall**
KPDT	14.1	23.4	11.5	15.4	6.1	20.8	13.7	Kappadam Tall
JMT	19.0	19.1	13.0	10.1	3.2	4.9	10.6	Jamaica Tall
SKGT	12.7	21.6	10.1	16.4	9.3	11.3	10.7	Sakhi Gopal Tall
WCT	**9.0**	**22.1**	**12.6**	**10.1**	**0.0**	**10.9**	**8.8**	**Indian West Coast Tall**
CALT	12.0	13.3	7.4	10.8	0.0	9.6	10.4	Indian West Coast Tall
CKT	22.9	21.3	11.8	14.6	14.4	4.9	8.7	Cameroon Kribi Tall
WAT	**24.8**	**32.5**	**19.6**	**18.5**	**14.1**	**0.0**	**15.9**	West African Tall
BRT	**17.8**	**42.4**	**10.4**	**19.4**	**33.5**	**39.0**	**0.0**	**Brasilian Tall**

The good representativeness achieved by the proposed procedure is illustrated by a factorial analysis of correspondences (FAC). The image obtained using this putative core collection is very similar to that obtained with all cultivars (see next page). The result is not significantly affected when the number of markers is reduced from 30 (as in the GCP study) to 13.

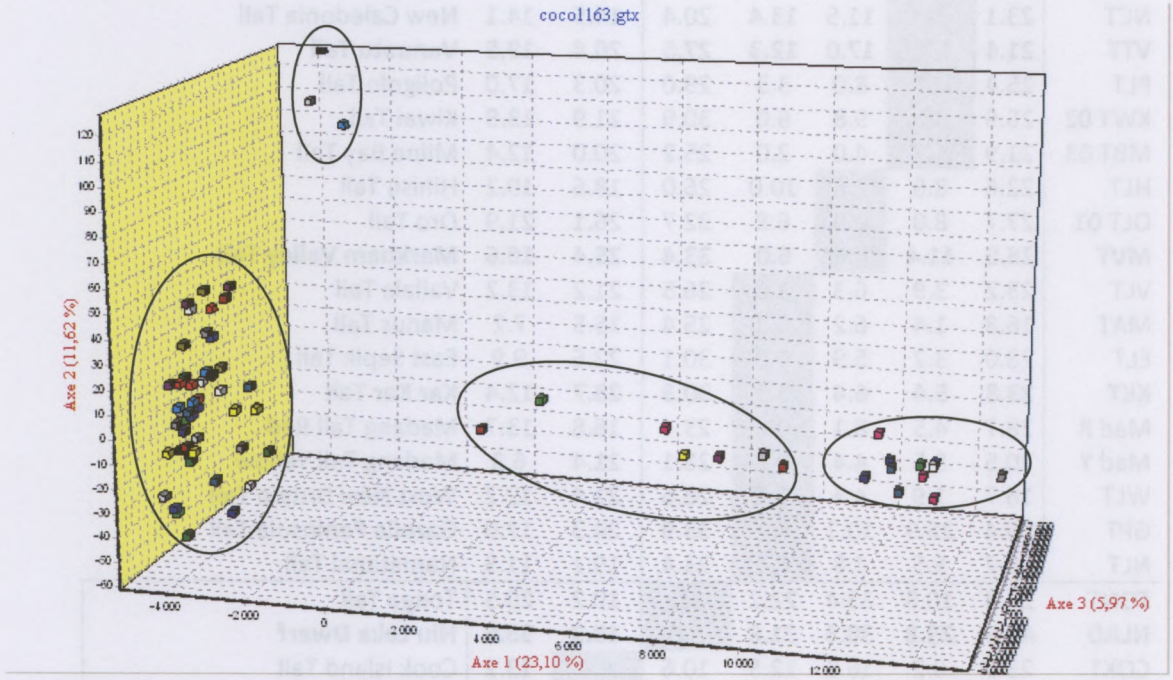


Figure 1: FAC on all populations

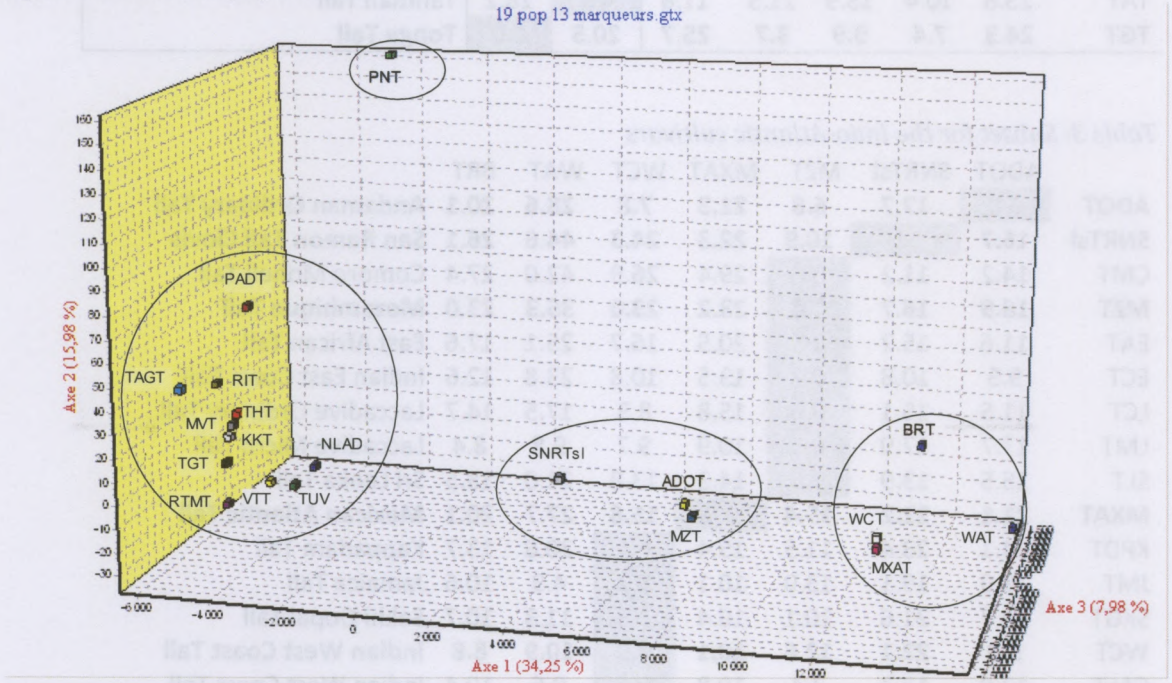


Figure 2: FAC on the Subsample

Conclusion:

The above procedure enabled us to extract a core subset of 19 coconut cultivars out of an initial 68 cultivar set (28% of the initial number). This subset takes into account the geographic structure of the species which was uncovered during systematic studies of the diversity of the species[6]. If we consider the South-Pacific region (Melanesia and Polynesia), the 7 populations RIT, VTT, MVT, KKT, NLAD, RTMT and TGT were representative of all the populations. A population such as BBRT was well represented by the VTT population. Similar results apply for the other two groups, for which only the following could be kept: 5 populations for the Panama, Micronesia and Southeast Asia group: PNT, TUV, PADT, TAGT and THT; 7 for the Indo-Atlantic group: ADOT, MZT, SNRTsI, MXAT, WCT, BRT and WAT. Once a sufficient number of loci is used, additional markers have little influence on the result.

It may be useful to propose an intuitive description of the procedure. We can describe the overall genetic diversity of a species in the form of an abstract multidimensional geometric space of finite "volume" (which, in turn can be figured imperfectly by a FAC). Each cross-pollinating population occupies a certain portion of this space, whose "volume" depends on genetic diversity in this population. In this description, a pure line or a clone would be represented by an almost dimensionless point and a highly diverse population would be represented by a large cloud. A population represents another one if its representative cloud covers the one of the represented population. The algorithm simply consists in selecting an OTU whose representative cloud is located in the "empty space" left by the already selected OTUs.

The important point in the above description is that once the genetic diversity of a species is adequately covered (with at least 19 coconut cultivars in our case), there is little benefit (from a conservation point of view) in increasing indefinitely the number of OTUs. In the case of coconut, a total of 60 cross-pollinating cultivars (the triple of what is strictly necessary) is largely sufficient to preserve diversity at a species level. This factor 3 can be justified to ensure desirable redundancy as noted above and because other considerations than molecular marker have to be taken into account. If each accession is kept on ½ hectare, this preservation programme demands an additional 60 hectare, which is reasonable. The preservation of rare alleles would be ensured in excellent conditions: at any locus, the preservation of alleles whose frequency is above 1‰ would be guaranteed.

This intuitive presentation can also be used as a guide for coconut Dwarf conservation which was not considered so far. While the genetic diversity of the Dwarfs, taken as a whole, is a small part of that of the species, each Dwarf cultivar covers only a very small portion of this diversity. As a result, the number of Dwarf cultivars worth conserving may be high, but correlatively, each of them can be represented by a small number of individuals (between 5 and 10). In conclusion, it would be reasonable to conserve each existing Dwarf cultivar provided it is really distinct. About 100 Dwarf cultivar names are listed of which an estimated 2/3 is really distinct and available. This would use less than 10 additional hectares.

It is important to note that the above discussion specifically addresses the needs for conservation for the purpose of preserving the genetic diversity of coconut. It is well understood that gene banks are also involved

---

[6] Gunn BF, Baudouin L, Olsen KM, 2011 Independent Origins of Cultivated Coconut (*Cocos nucifera* L.) in the Old World Tropics. *Plos One*,. 6(6): p. e21143.

in breeding and seed production activities and that, as such, they may wish to have more germplasm from a specific region or wish to establish a seed garden with a few specific varieties. Likewise, the purpose of this discussion is not to discourage the targeted collection of genotypes presenting specific and potentially useful traits.

Objective 2

« Identify gaps in coconut molecular characterization. The resulting list will include both cultivars/populations that were analyzed with an insufficient number of individuals and geographic locations of interest in the history of coconut evolution and where coconut diversity has not yet been characterized. »

Under this objective, we consider two cases:

- A number of cultivars or populations that were analyzed in the COGENT/BIOVERSITY/CIRAD analysis programs. The results are in the GCP and in the TropGENEdb databases. However, the number of individuals per cultivar was very low. This choice was made to be able to cover the global diversity of coconut at a reasonable cost. With the experience acquired. We propose to raise the figure at 10 or 15 per cultivar analysed in Talls and to 5 for Dwarfs.

- "Real" gaps, ie regions that are presently "white patches" on the map of coconut genetic diversity.

2-1 cultivars/populations that were analyzed with an insufficient number of individuals

Coconut genetic diversity has been studied based on a large number of populations represented by relatively small samples. Such a large sample, with more than 1300 individuals would never have been assembled without the help of the COGENT network and the result was an excellent coverage of most of the regions where coconut is cultivated, with small samples for each surveyed sample or collection accession. The typical sample size was only 5, although in a few cases, additional studies (eg. relative to Panama, Vanuatu, Seychelles Comoros and Madagascar Tall) resulted in larger samples. In the framework of a GCP study, a large number (30) of microsatellite markers was used. These data were analyzed recently[7] and the most important result was a strong differentiation of the coconut diversity in two geographic groups. Their areas of repartition are distinct except in a few regions: Madagascar, the Comoros, East Africa, Oman and the Andamans. The second outcome was a narrow connection between the diversity of coconut and human population movements such as migrations and intercultural contacts.

The following is an analysis of the CIRAD molecular data bases. Our suggestions to reinforce the sampling are given in the rightmost column. Our objective was to reach a total of 10 to 15 individuals in most Tall cultivars. A total of 5 being considered sufficient in the self-pollinating Dwarfs.

The columns are to be read as follows:

---

[7] Gunn BF, Baudouin L, Olsen KM, 2011 ref. cit.

- classification: the three-level classification key based on molecular and geographical basis
- Cult: cultivar abbreviation
- Cultname: cultivar name
- GCP: number of individuals analyzed with 30 markers
- KIT: number of individuals analyzed with 14 markers
- Total: sum of the GCP and KIT columns
- Add: suggested additions.

OTU means "operational taxonomic unit". It may be a cultivar or a population within a cultivar. Appendix 1 summarises the list of recommended accessions. A summary account of the analyses that were already performed is given in appendix 2 for collection accessions and in appendix 3 for populations sampled in situ.

Group A: Pacific coconuts (Talls and Dwarfs)

Comprises Tall and Dwarf cultivars from South-East Asia, Melanesia, Micronesia and Polynesia.

A1: South East Asian Dwarfs

Table 4: Malayan type Dwarfs (A1a)

classification	Cult	Cultname	GCP	KIT	Total	Add
A1a	CGD	Chowghat Green Dwarf		4	4	-
	GGBD	Gangabondam Green Dwarf		4	4	-
	MGD	Malayan Green Dwarf	5		5	-
	MIGD	Marshall Islands Green Dwarf	3	11	14	-
	MRD	Malayan Red Dwarf	2	9	11	-
	MYD	Malayan Yellow Dwarf	2	66	68	-
	PGD	Sri Lanka Green Dwarf	1	21	22	-
	RBD	Raja Brown Dwarf	3	9	12	-
	SKD	Salak Green Dwarf		5	5	-
Total A1a		*9 OTUs*	*16*	*129*	*145*	-

Table 5. Filipino type Dwarfs (A1b)

classification	Cult	cultname	GCP	KIT	Total	Add
A1b	AROD	Aromatic Green Dwarf	6	2	8	-
	BGD	Brazilian Green Dwarf		19	19	-
	CATD	Catigan Green Dwarf	2	4	6	-
	MGPD	Mangipod Green Dwarf		5	5	-
	PILD	Pilipog Green Dwarf	1	3	4	-
	TACD	Tacunan Green Dwarf	5		5	-
	THD	Thailand Green Dwarf	10		10	-
	XGD	Xiem Green Dwarf		2	2	3
Total A1b		*8 OTUs*	*24*	*35*	*59*	*3*
Total A1		**17 OTUs**	**40**	**164**	**204**	**3**

A2 : South Pacific Dwarfs

Note that this section includes only self-pollinating Dwarfs. The Niu Leka Dwarf is in group A6.

Table 6: South Pacific Dwarfs (A2)

classification	Cult	Cultname	GCP	KIT	Total	Add
A2	COD	Chowghat Orange Dwarf		5	5	-
	CRD	Cameroon Red Dwarf	2	3	5	-
	GPTr	Gazelle Red		4	4	-
	NICD	Nicaragua Green Dwarf		4	4	-
	PRD	Pemba Red Dwarf		6	6	-
	PYD	Papua New Guinea Yellow Dwarf		2	2	-
	RARD	Rabaul Red Dwarf		2	2	3
	RTB	King Coconut	2		2	5
	TRD	Tahiti Red Dwarf	1	4	5	-
	TRT	Talasea Red Tall		2	2	3
	VRD	Vanuatu Red Dwarf	1	4	5	-
Total A2		**11 OTUs**	**6**	**36**	**42**	**11**

A3: South-East Asian Talls

Table 7: Continental type Talls (A3a)

classification	Cult	cultname	GCP	KIT	Total	Add
A3a	DAUT	Dau Tall	6		6	9
	HAIT	Hainan Tall	7		7	8
	KAT	Cambodia Tall	10		10	-
	LPT	Lubuk Pakam Tall	5		5	5
	MLT	Malayan Tall	10		10	-
	SAT	Sawarna Tall	5		5	10
	SKT	Sarawak Tall	3		3	-
	TAAT	Ta Tall	7		7	-
	THT	Thailand Tall	18		18	-
Total A3a		*9 OTUs*	*71*		*71*	*32*

Table 8: Indonesian type Talls (A3b)

classification	Cult	Cultname	GCP	KIT	Total	Add
A3b	HK	Kopyor coconut	5		5	
	MYT	Mamuya Tall		5	5	
	PUT	Palu Tall	5		5	10
	TGT	Tenga Tall	5		5	10
	TKT	Takome Tall	5		5	10
Total A3b		*5 OTUs*	*20*	*5*	*25*	*30*

Table 9: Filipino type Talls (A3c)

classification	Cult	Cultname	GCP	KIT	Total	Add
A3c	BALT	Ballesteros Tall	7		7	
	BAYT	Baybay Tall	8		8	7
	MACT	Macapuno Tall	5		5	
	MXPT	Mexican Pacific Tall	39		39	
	PADT	Pandan Tall	6		6	9
	SNRT	San Ramon Tall	6		6	9
	TAGT	Tagnanan Tall	14		14	
Total A3c		*7 OTUs*	*85*		*85*	*25*
Total A3		**21 OTUs**	**176**	**5**	**181**	**87**

A4: Melanesian Talls

Table 10: North New Guinean type (A4a)

classification	Cult	Cultname	GCP	KIT	Total	Add
A4a	ELT	East Sepik Tall	8		8	7
	KKT	Karkar Tall	22	2	24	
	MDT_R	Madang Tall Red	4	1	5	
	MDT_Y	Madang Tall Yellow	4		4	
Total A4a		*4 OTUs*	*38*	*3*	*41*	*7*

Table 11: South New Guinean Type (A4b)

classification	Cult	Cultname	GCP	KIT	Total	Add
A4b	BBRT	Baibara Tall	5		5	5
	HLT	Hisihu Tall	5		5	5
	KWT	Kiwai Tall	4		4	5
	MBT	Milne Bay Tall	3	1	4	7
	OLT	Oro Tall	4		4	6
	PLT	Poligolo Tall	3		3	7
	VLT	Vailala Tall	10		10	7
Total A4b		*7 OTUs*	*34*	*1*	*35*	*42*

Table 12: PNG insular Type (A4c)

classification	Cult	Cultname	GCP	KIT	Total	Add
A4c	GPT	Gazelle Peninsula Tall	19	1	20	-
	GPTy	Gazelle Yellow	5		5	-
	MAT	Manus Tall	5		5	5
	NLT	Namatanai Tall	10		10	-
	WLT	West New Britain Tall	9	1	10	-
Total A4c		*5 OTUs*	*48*	*2*	*50*	*5*

Table 13: Markham Valley Tall (A4d)

classification	Cult	Cultname	GCP	KIT	Total	Add
A4d	MVT	Markham Valley Tall	20		20	-
Total A4d		*1 OTU*	*20*		*20*	-

Table 14: Vanuatu-Solomon type (A4e)

classification	Cult	Cultname	GCP	KIT	Total	Add
A4e	NCT	New Caledonia Tall	7		7	8
	RIT	Rennell Island Tall	12		12	3
	SIT	Solomon Island Tall	6		6	9
	VTT	Vanuatu Tall	335	100	435	
Total A4e		*4 OTUs*	*360*	*100*	*460*	*20*
Total A4		**23 OTUs**	**500**	**110**	**610**	**74**

A5: Micronesian Talls

Table 15: Micronesian Talls (A5)

classification	Cult	Cultname	GCP	KIT	Total	Add
A5	KIT	Kiribati Tall	21	4	25	-
	MIT	Marshall Island Tall	5		5	-[8]
	TUVT	Tuvalu Tall	17	1	18	-
Total A5		**3 OTUs**	**43**	**5**	**48**	**-**

A6: Polynesian Talls (includes the cross-pollinating Dwarf Niu Leka)

Table 16: Polynesian Talls (A6)

classification	Cult	Cultname	GCP	KIT	Total	Add
A6	COKT	Cook Island Tall	5	2	7	8
	NLAD	Niu Leka Dwarf	7		7	8
	NNMT	Niu Ni Magimagi Tall	4		4	11
	RTMT	Rotuman Tall	5		5	10
	TAT	Tahitian Tall	4	1	5	10
	TONT	Tonga Tall	5		5	10
Total A6		**6 OTUs**	**30**	**3**	**33**	**57**

A7: Panama Tall type

Table 17: Panama Tall type (A7)

classification	Cult	Cultname	GCP	KIT	Total	Add
A7	PET	Peru Tall		3	3	-
	PNT	Panama Tall	59	42	101	-
	PNT01	Panama Tall Aguadulce	26	1	27	-
	PNT03	Panama Tall Costa Rica	18	1	19	-
Total A7		**4 OTUs**	**103**	**47**	**150**	**-**
Total Group A		**85 OTUs**	**898**	**366**	**1264**	**232**

[8] Will be treated in question 2 of section 2.3

Group B: Indo-Atlantic Talls

B1: Typical Indo Atlantic Talls

Table 18: Typical Indo Atlantic Talls (B1)

classification	Cult	Cultname	GCP	KIT	Total	Add
B1	BRT	Brazilian Tall	72	97	169	
	CALT	Calangute Tall	3	2	5	10
	CKT	Cameroon Kribi Tall	5		5	5
	DRT	Dominican Republic Tall		96	96	
	ECT	Indian East Coast Tall	7		7	5
	JMT	Jamaica Tall	4	1	5	5
	LCT	Laccadive Ordinary Tall	3	2	5	5
	LMT	Laccadive Micro Tall	5		5	5
	MXAT	Mexican Atlantic Tall	9	2	11	
	SCT	Seychelles Tall		33	33	
	SKGT	Sakhi Gopal Tall	5		5	5
	SLT	Sri Lanka Tall	17	1	18	
	WAT	West African Tall	24		24	
	WCT	Indian West Coast Tall	6		6	5
Total B1		**14 OTUs**	**160**	**234**	**394**	**45**

B2: Introgressed Indo-Atlantic Talls

Table 19: Introgressed Indo-Atlantic Talls (B2)

classification	Cult	Cultname	GCP	KIT	Total	Add
B2	ADOT	Andaman Ordinary Tall	5		5	10
	CMRT	Comoros Tall		10	10	-
	CMT	Comoro Moheli Tall	7		7	-
	EAT	East African Tall	18		18	-
	KPDT	Kappadam Tall	6		6	9
	MDGT	Madagascar Tall		35	35	-
	MZT	Mozambique Tall	97	14	111	-
	SLT12	Sri Lanka Tall Margaret	2	1	3	12
	SNRT01	San Ramon Tall Clovis	5		5	10
Total B2		**9 OTUs**	**140**	**60**	**200**	**41**
Total Group B		**23 OTUs**	**300**	**294**	**594**	**86**
Grand Total		**107 OTUs**	**1198**	**660**	**1858**	**318**

Additional remarks: one of our conclusions of our report for consultance LOA APO/08/007 was that about 380 individuals from the Post-training studies conducted in the 2003/2005 period could be introduced into the online GeneTROP data base. This concerns the studies conducted in India, Sri Lanka, Mexico and probably Brazil. Actually, what is missing is the permission of the country and institution representatives. We also recommend coconut researchers to deposit their data into the GeneTROP database once results have been published.

2.2 Coconut growing regions that were never analyzed and deserve special attention

The fact that the microsatellite studies were launched within COGENT had many advantages but inevitably, non-member countries[9] are under-represented. We propose here a list of suggestions for further sampling. In this list, we tried to identify locations in order to answer a number of scientific questions. Our main goals are:

to look for new or as yet understudied diversity in coconut
to clarify our knowledge of the events which lead to the present-days diversity of coconut

We grouped our suggestions according to five scientific questions.

Question 1: where is the contact zone between the Indo-Atlantic and Pacific coconut groups?

The objective here is to explore the limit between the area of distributions of the Pacific and of the Indo-Atlantic groups. The remarkable difference between the Indo-Atlantic and Pacific groups strongly suggest that they have developed separately over a long period, which may have lasted at least until the end of the last glaciation. Since then, the limits may have fluctuated somewhat due to natural flotation and to human population movements and trade. Locating the contact zone is interesting both for anthropological[10] and agronomic reasons. In fact, contact regions are hot spots of genetic diversity and are of particular interest for plant breeders. The intended survey should cover the north-western coast of Thailand, Myanmar and Bangladesh. Pakistan and the Maldives, which haven't yet been analyzed, are added for comparison.

Table 20: Proposed analyses in the Indian Ocean

Country	Sites	Note
Thailand (Indian Ocean coast)	3	
Myanmar	5	
Bangladesh	5	Collect from COGENT genebank
Pakistan	5	Collect from COGENT genebank
Maldives	4	
Total	22	

Question 2: what is the contribution of early Austronesian migrations to the dissemination of coconut?

The common point of the islands cited here is that, presently, their populations is mainly Polynesian. However, it is highly suspected that their initial settlement pre-dated the arrival of Polynesians. Little is known about the original populations. Material evidence is scarce and often indirect (e.g. change in flora). Coconut could help us learn more about them. Conversely, chances of identifying as yet unrecognized diversity in coconut are substantial.

[9] and those which were not member when the analyses were performed.
[10] because coconut is a marker of human activities and exchanges.

Palau was settled between the 4th and 3rd millennia BP[11], most likely from the Philippines or Indonesia. This was before the Polynesian expansion and Palauan, a Malayo-Polynesian language doesn't belong to the Oceanic family, which comprises the Polynesian languages[12].

If we move eastward, we meet the Marianas Islands. Again, a non-Oceanic language (Chamorro) is found in Guam. The traditional houses recall South-East Asian architecture. Disturbance indicators (charcoal, pollen of different taxa) indicate that the first settlements occurred about 5000 Y BP[13]. It is said that coconut varieties similar to the San Ramon exist in Guam. According to the archaeologist Davison[14] about Nukuoro, " the Polynesian languages now spoken and, in some cases, the origin traditions of the present inhabitants, suggest that effective Polynesian colonization took place relatively recently; often long after initial human colonization of the islands in question, as identified by archaeology."

Marshall Islands are said to have been settled in the second millennium BC. artifacts (wood) dated 3500, 2800 and 2600 BP were found on the Bikini atoll[15].

Table 21: Proposed analyses in the northern Pacific Ocean

Island/archipelago	Sites	Notes
Palau	2	
Marianas Isl.	3	
Marshall Isl.	3	
Total	8	

Question 3: what is the pattern of diffusion of coconut from Central Polynesia to Hawai'i?

Considering that French Polynesia is located at the center of the "polynesian triangle" it appears to be seriously underrepresented in coconut molecular studies. Society Islands and Tuamotus are represented by only a small number of individuals and neither Austral, Gambier or Marquesas Islands have been sampled. The Marquesas are considered as the source of the settlers of Hawai'i and the coconut palms of both archipelagos are probably closely related. Molecular studies can help confirming this hypothesis and assessing the presence of a possible population bottleneck.

---

[11] Welch DJ, 2002 Archaeological and paleoenvironmental evidence of early settlement in Palau. *Indo-Pacific prehistory association bulletin* 22 (Melaka papers vol. 6) 161-173

[12] Lewis P. Ed., 2009.Ethnologue, Languages of the world. International library of Congress.

[13] Athens JS, Dega M, Ward JV 2004. Austronesian colonization of the Mariana Islands: the paleoenvironmental evidence. *Indo-Pacific prehistory association bulletin* 24 (Taipei papers vol. 2) 21-30

[14] J Davison 1992. New evidence about the date of colonization of Nukuoro Atoll, a Polynesan outlier in the Eastern Caroline Islands. Journal of the Polynesian Society. 101, n° 3

[15] Streck CF, 1990. Prehistoric settlement in Eastern micronesia : Archaeology on Bikini Atoll, Republic of the Marshall Islands.

Table 22: Proposed analyses in Eastern Polynesia

Archipelago	Sites	Notes
Society	3	
Tuamotus	3	
Austral	4	
Gambier	4	
Marquesas	4	
Hawai'i, Maui, Molika'i	3	
Midway Isl.	3	
Total	24	

Question 4: what are the consequences of the Arabic and German influence on the coasts of Tanzania?

The Indo-Atlantic alleles are predominant in the genomes of coconut in Tanzania. However, about 1/3 of the alleles are from the Pacific group. The only African cultivars which exhibited some degree of resistance to a Lethal-Yellowing-type diseased caused by Phytoplasmas are from northern Tanzania. Although this resistance is not complete, the phenomenon is interesting because all other resistance sources are from the Pacific group. Resistance in theos East African populations could thus be a consequence of the introgressed status of the Tanzanian germplasm. A particularity of Tanzania is that there were two waves of intensification of coconut cultivation. The first one occurred in the North, under an Arabic influence in the 19th Century. The second one occurred in the South, under a German influence at the beginning of the 20th Century. Our objective is to assess the difference between Northern and Southern planting materials, in terms of diversity and of allelic composition.

Twelve collection sites, located from the North to the South of the country should be sampled. Both healthy and diseased palms should be sampled. a study of the diversity of the pathogen should be conducted in parallel.

Table 23: Proposed analyses in Eastern Polynesia

Archipelago	Sites	Notes
Northern region	3	
Tuamotus	3	
Austral	4	
Gambier	4	
Marquesas	4	
Hawai'i, Maui, Molika'i	3	
Midway Isl.	3	
Total	24	

Question 5: what was the extent of the distribution of the Pre-Columbian coconuts on the Pacific coast of America?

Although Cocos nucifera has distant ancestors in South America, it is agreed that its lineage became extinct on the continent. But, when the Spanish explorers crossed the Isthmus of Panama, they observed it growing spontaneously at several locations[16]. Our aim is to refine the scenario of this introduction. Coconuts

[16] Patiño VM. 2002. Historia y Dispersión de los frutales nativos del Neotrópico. Publication CIAT n° 368. pp 241 to 270.

populations located close to two historical sites were abundantly sampled. However, other sites haven't yet received much attention. This concerns the Pearl Islands archipelago where coconut was first mentioned, the Isla del coco where it is said to grow by the sea (this excludes a confusion with the superficially similar species *Euterpe macrospadix*). Sites worth being explored in the region extending from Choco to the North of Peru include the Bahia Candelaria, Cabo Corrientes, Isla Gorgone, Rio Baudo, Tumaco, Bahia de Caraquez, Gulf of Guayaquil., Valdivia, Tumbes and Sullana. We do not propose to samples northern to Costa Rica because we understand that an interesting study of the Pacific coast of America, from Mexico to Costa Rica is under way.

Table 23: Analyses on the Pacific coasts of America

Region/island	Sites	Notes
Pearl Islands archipelago	3	
Isla del coco	2	One site in Wafer Bay, the other one in an isolated bay on the West coast
Bahia Candelaria,	3	
Cabo Corrientes,	1	
Isla Gorgone	1	
Rio Baudo	1	
Tumaco,	1	
Bahia de Caraquez	3	
Gulf of Guayaquil	3	
Valdivia,	1	
Tumbes	1	
Sullana	1	Already three palms analyzed
	21	

Question 6: what are the common traits of spontaneous or sub-spontaneous populations:

Populations from East Samar in the Philippines (Bukidnon?) Cocos-Keeling. Up to 5 more populations proposed by specialists from South India, SEA and South Pacific

This list is not limitative. A number of isolated places also deserve consideration: Tokelau and Wallis and Futuna, Palmyra etc in the Pacific.

Table 24: Analyses in potentially wild populations

Region	Sites	Notes
East Samar	2	
Cocos Keeling	2	
Lizard Island	1	
India, Sri Lanka	3	Proposed by country member researchers
SEA	3	Proposed by country member researchers
SP	3	Proposed by country member researchers
Total	11	

Collection procedure: In all cases, it is recommended to collect from a local population (no record of importation from a foreign country). The age of the palms is indifferent if the coconut population is

---

Baudouin, L., B.F. Gunn, and K.M. Olsen. 2014. The presence of coconut in southern Panama in pre-Columbian times: clearing up the confusion. *Annals of Botany* 113: 1-5.

homogenous. Otherwise, preference should be given to old palms. Sampling will be made according to the recommended practices. It will be accompanied by a brief description of the site. If possible, photograph of the plant and of the fruits (according to a procedure defined by R. Bourdeix) can be added. We recommend to sample DNA from 10 palms per site. Given that the total number of sites is 96, it would represent 960 individuals. However, we consider this as a long term objective. We would consider it as a success if two of the proposed programs are completed in the next 5 years.

Objective 3:

"Interact and communicate with the International Thematic Action Group on Coconut Genomics and obtain a consensus about points 1 and 2, in order to include the conclusions of these studies in the Global Strategy for Conservation and Use of Coconut Genetic Resources to be published in December 2013."

The coconut genomics ITAG (International Technical Action Group) was consulted and a draft of the present report was transmited. No negative response was received. Prof. Heslop-Harrison responded and provided useful advice, which wad taken into account. See appendix 6.

He put forward an interesting point: the microsatellite kit will not last forever. It could be superseded in 2 to 3 years. Sequencing efforts are already producing more microsatellite markers[17], they will also open the way to high throughput genotyping methods such as DNA chips and genotyping buy sequencing. eventually, resequencing could be considered the most efficient approach. This does not mean that the kit will suddenly lose its interest and most of of the projects mentioned in this report will retain their attractivity. However, they will probably be the last large scale projects conducted with the kit. In addition, once the gap filling operations are achieved, we will have learned most of the knowledge they can offer in terms of population genetics.

Inevitably, new techniques will pave the way to more efficient approaches including genomic selection, GWAS (genome wide association studies). there will be a need to connect these approach to the results obtained with the kit. This can be done by including a sub-sample of well chosen genotypes into the studies conducted with the new techniques, just in the same way as was already done when movinf from RFLPand AFLP to SSRs[18].

Objective 4:

"Assist the Philippines and Sri Lanka in developing their long term ex situ conservation policy of local germplasm. This assistance will consist in using molecular markers as a way of reducing redundancy in the collection while ensuring that genetic diversity at the country level is appropriately represented."

---

[17] Yong, X., L. Yi, Y. Yaodong, F. Haikuo, X. Wei, A.S. Mason, Songlin Zhao, R. Sager, and F. Qiao. 2013. Development of microsatellite in *Cocos nucifera* and their application in evaluating the level of genetic diversity of *Cocos nucifera*. *plant Omics Journal* **6:** 193-200.

[18] Teulat, B., C. Aldam, R. Trehin, P. Lebrun, J.H.A. Barker, G.M. Arnold, A. Karp, L. Baudouin, and F. Rognon. 2000. An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. *Theoretical and Applied Genetics* **100:** 764-771.

Outline of country-level analysis programmes in Sri Lanka and in the Philippines[19]

The objective of the program we propose in Sri Lanka and in the Philippines is to evaluate the structure of genetic diversity of Tall coconuts at the country level. Genetic parameters such as inbreeding coefficients, population level diversity and between population differentiation will be evaluated and used as guidelines to estimate the number of accessions and the accession sizes needed to adequately conserve diversity at the country level. In addition, the comparison between Sri Lanka and the Philippines will be especially enlightening because they harbor two different sub-species of coconut. We do not advise to attempt a systematic coverage of all populations that were sampled. This would not be realistic given that, to be significant, the parameters must be evaluated on a sufficient number of individuals. We estimate that this number should not be less than 20. Taking this into account, we consider that the number of OTUs (here, local cultivar or population[20]) should be comprised between 10 and 15 per country. (200 to 300 individuals per country)

A secondary objective would be to encourage a comparison between microsatellite marker sets: it happens that both Sri Lanka and the Philippines have been pioneers in microsatellite analyses. We propose to combine several marker kits (eg 6 markers developed at CIRAD and 4 from Sri Lanka and from the Philippines. The analyses should be done in the same lab (possibly using GCP genotyping facilities) or, failing this, at least 10 individuals should be analyzed in common by both laboratories.

The choice of the OTUs has to be made by the national researchers based on geographic and phenotypic criteria (trying to include as much diversity as possible).

Objective 5:

"Invite India, the Philippines, Mexico, Brazil[21], China and Sri Lanka to contribute data to TropgenDB. As regards genotyping data, those obtained with the coconut kit developed by CIRAD should be considered in priority but other data is also welcome. Markers, linkage maps and QTL data are equally welcome."

The permission to use genetyping data obtained during the post-training experiment(2005-2006) was solicited. Introduction of further data (genotyping or marker data) was suggested.

RV Manimekelai, Rk Rajesh and V. Arunachalam (CPCRI, India)
R Rivera (PCA, the Philippines),
D. Zizumbo (CICY, Mexico)
Y Yaodong (CRI-CATAS, China)
C Perera and L Perera (CRI, Sri Lanka)

We received the permission from CPCRI to include data on 152 individuals from the pos-training study. CPCRI also envisages the inclusion of further data in the future. We expect more responses in a next future.

---

[19] this study could be extended to India and Indonesia if these countries are interested.
[20] in the nomenclature of coconut diversity, "populations" are the level below cultivar.
[21] not done (the data exist already in TropGENE-db)

Objective 6:

Make the user-interface of the coconut section of TropgenDB more accessible and complete for international users. This multi-crop TropgenDB WEB database hosted by CIRAD holds molecular data of 1321 coconut individuals and is accessible online to the whole coconut research community. The objective is to add more data and standardized passport information (the conservation site code and unique access number, international cultivar name and population number) for a better identification of the individuals.

Thank to the contribution of Chantal Hamelin, the interface of the coconut section of TropeGENE-db was improved:

"Studies" subsection: the description of the studies is more informative. Whenever applicable, the text of the relevant publication can be accessed on click.

"Genotypes" subsection: the requests are more convivial, the columns are more informative. Data were added: 647 individuals studied with 13 markers genotyped by CIRAD.

"Markers" subsection: the order of the column was cnanged for better readability. Whenever applicable, the text of the relevant publication can be accessed on click.

"QTLs" subsection: unchanged.

More improvement is planned for year 2014 to make the database more attractive for visitors as well as for potential contributors.

Appendix 1: List of suggested accessions for complementary analyses

Genetic group	Abbréviation	Cultivar/ population	Done	To do	Genebank	Accession
A1a	XGD	Xiem Green Dwarf	2	3	DGEC	DGEC/L007
A2	RARD	Rabaul Red Dwarf	2	3	SRS	SRS RARD
A2	RTB	King Coconut	2	5	CRI	CRI RTB01 I2
A2	TRT	Talasea Red Tall	2	3	SRS	SRS TRT
A3a	DAUT	Dau Tall	6	9	DGEC	DGEC/L020
A3a	HAIT	Hainan Tall	7	8	CHIN	CATAS 2
A3a	LPT	Lubuk Pakam Tall	5	5	IPRI	IPR-MAT001
A3b	SAT	Sawarna Tall	5	10	IPRI	IPR-MAT002R1
A3b	PUT	Palu Tall	5	10	IPRI	IPR-MAT006R1
A3b	TGT	Tenga Tall	5	10	IPRI	IPR-MAT007R2
A3b	TKT	Takome Tall	5	10	IPRI	IPR-MAT025R2
A3c	BAYT	Baybay Tall	8	7	PHL	PPC BAYT A09
A3c	PADT	Pandan Tall	6	9	PHL	ZRC 043-24
A3c	SNRT	San Ramon Tall	6	9	PHL	PPC SNRT A15
A4a	ELT	East Sepik Tall	8	7	SRS	SRS ELT04
A4b	BBRT	Baibara Tall	5	5	SRS	SRS BBR
A4b	HLT	Hisihu Tall	5	5	SRS	SRS HLT
A4b	KWT	Kiwai Tall	4	5	SRS	SRS KWT01
A4b	MBT	Milne Bay Tall	4	7	SRS	SRS MBT03
A4b	OLT	Oro Tall	4	6	SRS	SRS OLT01
A4b	PLT	Poligolo Tall	3	7	SRS	SRS PLT
A4b	VLT	Vailala Tall	10	7	SRS	SRS VLT02
A4c	MAT	Manus Tall	3	5	SRS	SRS MAT01
A4e	NCT	New Caledonia Tall	7	8	VT	VT GNC
A4e	RIT	Rennell Island Tall	12	3	SRS	SRS RLT
A4e	SIT	Solomon Island Tall	6	9	VT	VT GSN1 R1
A6	COKT	Cook Island Tall	7	8	local source	
A6	NLAD	Niu Leka Dwarf	7	8	WS	WS006
A6	NNMT	Niu Ni Magimagi Tall	4	11	local source	
A6	RTMT	Rotuman Tall	5	10	TCC	TCC ROT
A6	TAT	Tahitian Tall	5	10	SMD	SMD GPY1 R1
A6	TONT	Tonga Tall	5	10	TON	TON 6

Genetic group	Abbréviation	Cultivar/ population	Done	To do	Genebank	Accession
B1	CALT	Calangute Tall	5	10	IND	IND071 R1
B1	CKT	Cameroon Kribi Tall	5	5	SMD	SMD GCA R1
B1	ECT	Indian East Coast Tall	7	5	IND	IND127
B1	JMT	Jamaica Tall	5	5	CIB	CIB JMT R1
B1	LCT	Laccadive Ordinary Tall	5	5	IND	IND008 R1
B1	LMT	Laccadive Micro Tall	5	5	IND	IND030 R2
B1	SKGT	Sakhi Gopal Tall	5	5	IND	IND041 R1
B1	WCT	Indian West Coast Tall	6	5	IND	IND069 R2
B2	ADOT	Andaman Ordinary Tall	5	10	IND	IND018 R2
B2	KPDT	Kappadam Tall	6	9	IND	IND001 R1
B2	SLT12	Sri Lanka Tall Margaret	3	12	CRI	CRI SLT12
B2	SNRT01	San Ramon Tall Clovis	5	10	CRI	CRI SNRT01R1

Appendix 2: List of accessions already analyzed with the standard kit

Cult	cultname	CGRD Accession	Number
HAIT	Hainan Tall	CATAS 5	7
JMT	Jamaica Tall	CIB JMT	5
PET	Peru Tall	CIB PET	3
SKT	Sarawak Tall	CIB SKT	3
MXAT	Mexican Atlantic Tall	CICY MXAT01	11
MXPT	Mexican Pacific Tall	CICY MXPT01	12
MXPT	Mexican Pacific Tall	CICY MXPT02	14
MXPT	Mexican Pacific Tall	CICY MXPT03	13
RTB	King Coconut	CRI RTB 01	2
SLT12	Sri Lanka Tall Margaret	CRI SLT 12	3
SLT	Sri Lanka Tall	CRI SLT 27	3
SLT	Sri Lanka Tall	CRI SLT02 R3	2
SNRT01	San Ramon Tall Clovis	CRI SNRT 01	5
XGD	Xiem Green Dwarf	DGEC/L007	2
TAAT	Ta Tall	DGEC/L009	7
DAUT	Dau Tall	DGEC/L020	6
KPDT	Kappadam Tall	IND 001	1
GGBD	Gangabondam Green Dwarf	IND003 R1	4
LCT	Laccadive Ordinary Tall	IND008 R1	5
SKGT	Sakhi Gopal Tall	IND041	5
EAT	East African Tall	IND042	4
WCT	Indian West Coast Tall	IND069	6
CALT	Calangute Tall	IND073	5
ECT	Indian East Coast Tall	IND127	7
SKD	Salak Green Dwarf	IPR-MAD020R1	5
MYT	Mamuya Tall	IPR-MAT038	5
RBD	Raja Brown Dwarf	MAP D 6	2
LPT	Lubuk Pakam Tall	MAP T 1	5
SAT	Sawarna Tall	MAP T 2	5
PRD	Pemba Red Dwarf	NCDP-D1 M R1	6
EAT	East African Tall	NCDP-T1B CHA	4
EAT	East African Tall	NCDP-T1H LBS	5
EAT	East African Tall	NCDP-T1K KIM	5
CKT	Cameroon Kribi Tall	SMD GCA R1	5
KAT	Cambodia Tall	SMD GCB10 R1	5
KAT	Cambodia Tall	SMD GCB7 R1	5
CMT	Comoro Moheli Tall	SMD GCO	7
TKT	Takome Tall	SMD GDO1	5
TGT	Tenga Tall	SMD GDO2	5
PUT	Palu Tall	SMD GDO3	5
MLT	Malayan Tall	SMD GML R1	10

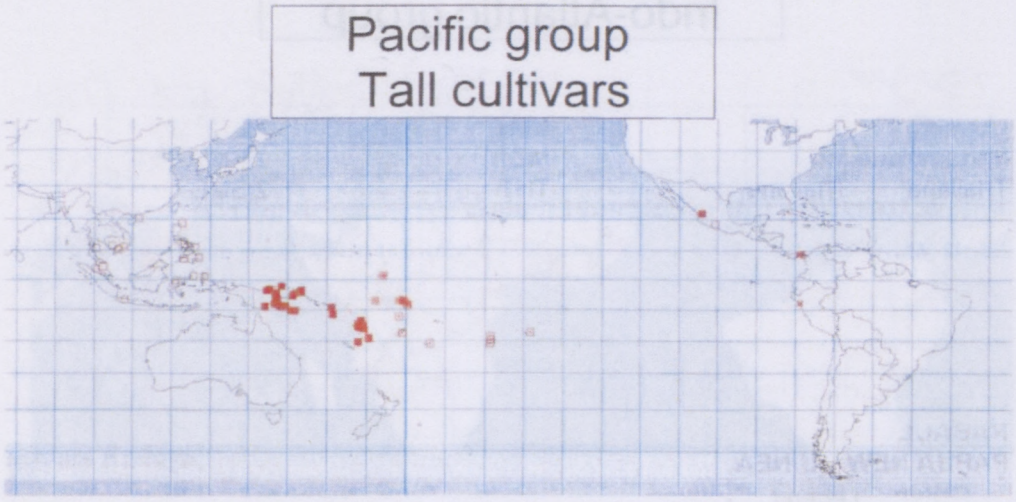
Cult	cultname	CGRD Accession	Number
MZT	Mozambique Tall	SMD GMZ	7
ADOT	Andaman Ordinary Tall	SMD GND2 R1	5
KPDT	Kappadam Tall	SMD GND5 R1	5
LMT	Laccadive Micro Tall	SMD GND7	5
KKT	Karkar Tall	SMD GNG1 B	5
MVT	Markham Valley Tall	SMD GNG3	5
GPT	Gazelle Peninsula Tall	SMD GNG4	5
WAT	West African Tall	SMD GOA4 R1	10
WAT	West African Tall	SMD GOA6 R	10
PNT01	Panama Tall Aguadulce	SMD GPA1	14
PNT	Panama Tall	SMD GPA2	23
BAYT	Baybay Tall	SMD GPH4	5
TAT	Tahitian Tall	SMD GPY1 R1	5
RIT	Rennell Island Tall	SMD GRL	7
RTMT	Rotuman Tall	SMD GRT	5
SLT	Sri Lanka Tall	SMD GSL	13
SIT	Solomon Island Tall	SMD GSN	1
SIT	Solomon Island Tall	SMD GSN R1	5
TONT	Tonga Tall	SMD GTG	5
THT	Thailand Tall	SMD GTH 1	7
TAGT	Tagnanan Tall	SMD GTN A	5
TAGT	Tagnanan Tall	SMD GTN B	8
VTT	Vanuatu Tall	SMD GVT	14
RBD	Raja Brown Dwarf	SMD NBN	5
RBD	Raja Brown Dwarf	SMD NDO	5
MYD	Malayan Yellow Dwarf	SMD NJG R2	7
MYD	Malayan Yellow Dwarf	SMD NJM R1	2
NLAD	Niu Leka Dwarf	SMD NNL R1	7
CRD	Cameroon Red Dwarf	SMD NRC R2	5
MRD	Malayan Red Dwarf	SMD NRM R1	5
TRD	Tahiti Red Dwarf	SMD NRY	5
BGD	Brazilian Green Dwarf	SMD NVE R1	15
MGD	Malayan Green Dwarf	SMD NVM R1	5
CATD	Catigan Green Dwarf	SMD NVP2	5
TACD	Tacunan Green Dwarf	SMD NVP3	5
PILD	Pilipog Green Dwarf	SMD NVP5	3
AROD	Aromatic Green Dwarf	SMD NVP7	6
PGD	Sri Lanka Green Dwarf	SMD NVS B	5
THD	Thailand Green Dwarf	SMD NVT	10
BBRT	Baibara Tall	SRS BBRT	5
ELT	East Sepik Tall	SRS ELT03	3
ELT	East Sepik Tall	SRS ELT04	5
GPT	Gazelle Peninsula Tall	SRS GLT01	5
GPT	Gazelle Peninsula Tall	SRS GLT04	5

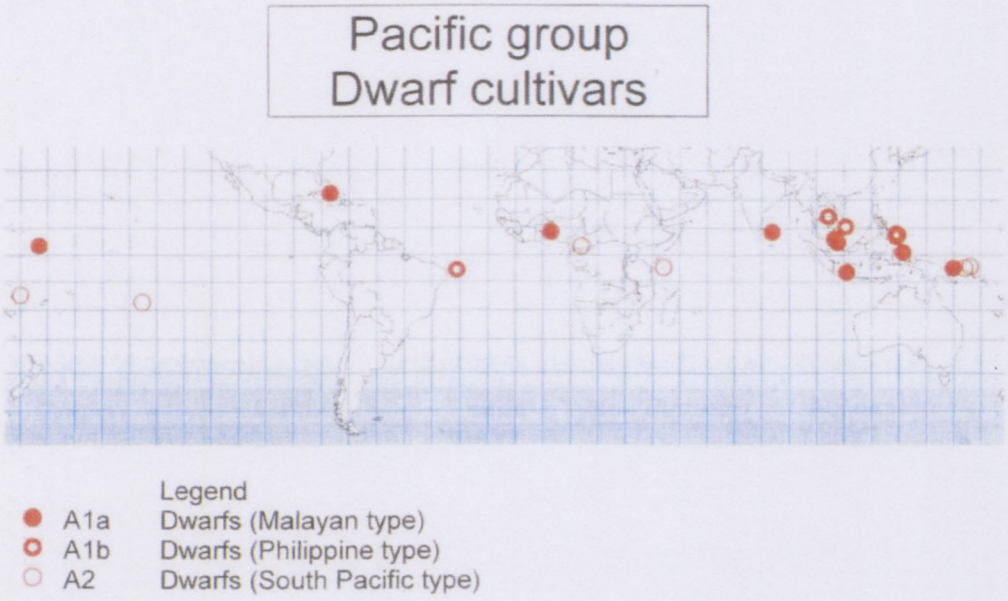
Cult	cultname	CGRD Accession	Number
GPT	Gazelle Peninsula Tall	SRS GMT05	5
GPTr	Gazelle Red	SRS GRT02	4
HLT	Hisihu Tall	SRS HLT	5
VLT	Vailala Tall	SRS ILT	5
KKT	Karkar Tall	SRS KKT01	5
KKT	Karkar Tall	SRS KKT02	9
KKT	Karkar Tall	SRS KKT03	5
KWT	Kiwai Tall	SRS KWT01	4
MAT	Manus Tall	SRS MAT01	2
MAT	Manus Tall	SRS MAT02	3
MBT	Milne Bay Tall	SRS MBT03	4
MVT	Markham Valley Tall	SRS MVT01	10
MVT	Markham Valley Tall	SRS MVT02	5
NLT	Namatanai Tall	SRS NLT01	5
NLT	Namatanai Tall	SRS NLT04	5
OLT	Oro Tall	SRS OLT01	4
PLT	Poligolo Tall	SRS PLT	3
PYD	Papua New Guinea Yellow Dwarf	SRS PYD	2
RARD	Rabaul Red Dwarf	SRS RARD	2
TRT	Talasea Red Tall	SRS TRT	2
VLT	Vailala Tall	SRS VLT02	5
WLT	West New Britain Tall	SRS WLT01	5
WLT	West New Britain Tall	SRS WLT02	5
BAYT	Baybay Tall	Unknown accession	1
CATD	Catigan Green Dwarf	Unknown accession	1
MACT	Macapuno Tall	Unknown accession	5
MRD	Malayan Red Dwarf	Unknown accession	1
PILD	Pilipog Green Dwarf	Unknown accession	1
SNRT	San Ramon Tall	Unknown accession	1
TAGT	Tagnanan Tall	Unknown accession	1
RIT	Rennell Island Tall	VT GRL1 R3	5
NCT	New Caledonia Tall	VT NCT	7
MRD	Malayan Red Dwarf	VT NRM R1	5
VRD	Vanuatu Red Dwarf	VT NRV R2	5
VTT	Vanuatu Tall	VT TAN 2	2
BALT	Ballesteros Tall	ZRC 027 7B	7
PADT	Pandan Tall	ZRC 043-24	6
BAYT	Baybay Tall	ZRC BAYT	2
AROD	Aromatic Green Dwarf	ZRC FD3	2
MGPD	Mangipod Green Dwarf	ZRC PD18	5
SNRT	San Ramon Tall	ZRC PT 4	5

Appendix 3: List of analyzes made on in situ populations, grouped by cultivar.

Cult	cultname	CGRD Accession	Number
BGD	Brazilian Green Dwarf	Not an accession	4
BRT	Brazilian Tall	Not an accession	169
CGD	Chowghat Green Dwarf	Not an accession	4
CMRT	Comoros Tall	Not an accession	10
COD	Chowghat Orange Dwarf	Not an accession	5
COKT	Cook Island Tall	Not an accession	7
DRT	Dominican Republic Tall	Not an accession	96
GPTy	Gazelle Yellow	Not an accession	5
HK	Kopyor coconut	Not an accession	5
KIT	Kiribati Tall	Not an accession	25
MDGT	Madagascar Tall	Not an accession	35
MDT_R	Madang Tall Red	Not an accession	5
MDT_Y	Madang Tall Yellow	Not an accession	4
MIGD	Marshall Islands Green Dwarf	Not an accession	14
MIT	Marshall Island Tall	Not an accession	5
MYD	Malayan Yellow Dwarf	Not an accession	59
MZT	Mozambique Tall	Not an accession	104
NICD	Nicaragua Green Dwarf	Not an accession	4
NKFT	Tall type niu Kafa	Not an accession	1
NNMT	Niu Ni Magimagi Tall	Not an accession	4
PGD	Sri Lanka Green Dwarf	Not an accession	17
PNT	Panama Tall	Not an accession	78
PNT01	Panama Tall Aguadulce	Not an accession	13
PNT03	Panama Tall Costa Rica	Not an accession	19
SCT	Seychelles Tall	Not an accession	33
THT	Thailand Tall	Not an accession	11
TUVT	Tuvalu Tall	Not an accession	18
VTT	Vanuatu Tall	Not an accession	419
WAT	West African Tall	Not an accession	4

Appendix 4: Places of origin of the sampled cultivars





Indo-Atlantic group

Legend

◆	B1	Indo-Atlantic Tall (no introgression)
◈	B2	Indo-Atlantic Tall (low introgression rate)
◇	B3	Indo-Atlantic Tall (moderate to high introgression rate)

Appendix 5: the procedure to select a representative subset

Step 1: After performing molecular analyses, create an appropriate entry file for Geneclass 2 (see manual).

Step 2: perform self-assignment using geneclass 2 with the following options:

computation goal: application = "assign/exclude population as origin of individuals" and assign = "individuals".

Criteria for computation: "Bayesian methods, Rannala &Mountain"

Probability computation: do *not* mark "Enable probability computation"

Important: leave the "Samples to be assigned" field blank. Geneclass 2 will automatically apply the "leave one out" procedure.

Step 3: go to tab "Results" and export the output table as a .CSV file

Step 4: use R software and run the four functions below.

Function	Parameters	Output	Comment
read.score	The name of the Geneclass 2 output file	The individual assignment score matrix	This is a rectangular matrix
aver.score	The individual assignment score matrix	The population assignment scores	this is a square matrix (one row per population)
init.set	- The population assignment scores - The names of the populations initially chosen - The names of the populations initially excluded	a table with the populations ranked by ascending "representation scores".	the populations are assigned a status ("in", "candidate" or "out")
accept	- the population assignment scores - the name of a new member (optional)	The same table, updated	by default, the "candidate" population with the highest representation score is added. However, any candidate population can be added instead.

The "accept" function is to be repeated until one of the following occurs:

a preset number of populations was chosen

the highest representation score is below a fixed number

the procedure tends to choose populations that are known to be very close to already chosen populations.

Code

```
read.score<-function(filename){
# This function takes a file name, which must be
# an individual assignment output file frome Geneclass 2
# It extracts the assignment score matrix
u<-scan(filename,"", skip=10,sep=";",quote=NULL)
index<-grep("/",u) # A étrick" to locate tne beginning of a row
maxw<-length(u)
start<-min(grep("(rang)|(rank)", u))+15        #Geneclass 2 is bilingual.
```

```
end<-grep("(Nb. of loci)|Nb. de locus)",u)-1  #Geneclass 2 is bilingual.
pop<-u[start:end]
np<-length(pop)
#one row
out<-NULL
ind<-NULL
ind<-end+np-12
      for(i in index){
      ind<-c(ind,u[i])
      startl<-i+16
      endl<-startl+np-1
      out<-rbind(out,as.numeric(u[startl:endl]))
      }
rownames(out)<-u[index]
colnames(out)<-pop
as.data.frame(out)
}
```

```
aver.score<-function(df){
#Starting from an individual score matrix, obtains a square average score matrix
#implicitly, it was a self-assignment procedure
#Thus the result is a square matrix
colsplit<-strsplit(rownames(df),"/")
pop<-as.factor(sapply(colsplit, function(x) x[1]))
out<-by(df,pop,mean)
out<-t(sapply(out,rbind))
out<-t(apply(out,1, function(x) x-min(x)))
rownames(out)<-levels(pop)
colnames(out)<-colnames(df)
out
}
```

```
init.set<-function(table, pop.in=NULL, pop.out=NULL){
#Takes the asssignment scores and population names
#Creates a table with the population names,
#the status (in, out and candidate)
#and the inclusion score
pop<-colnames(table)
status<-rep("candidate", length(pop))
status[pop %in%  pop.in]<-"in"          # To be included
status[pop %in%  pop.out]<-"out" # Not available
stable<-table[,status=="in"]
scoremin<-switch(class(stable),matrix=apply(stable,1, min),stable)
out<-data.frame(population=pop, status, low.score=scoremin)
levels(out$status)<-union(levels(out$status),"out")
out[order(out[3]),]
}
```

```
accept<-function(table, sets, pop.in=NULL){
#Takes the asssignment scores, a previously calculated set table and population names
#updates the set table
if(is.null(pop.in)) {
      high.score<-max(sets$low.score[sets$status=="candidate"])
      pop.in<-sets$population[sets$low.score==high.score]
      }
else {
      if(length(pop.in)!=1) stop("only one population at a time, please")
      if(!pop.in %in% sets$population[sets$status=="candidate"]) stop("This is not a
candidate population")
      }
 sets$status[sets$population %in%  pop.in]<-"in"
 stable<-table[,sets$status=="in"]
 scoremin<-switch(class(stable),matrix=apply(stable,1, min),stable)
 sets$low.score<-scoremin
 sets[order(sets$low.score),]
}
```

Appendix 6: Message from P. Heslop-Harrison
(with answers by L. Baudouin in red).

De : Luc Baudouin [luc.baudouin@cirad.fr]
Date d'envoi : mercredi 13 novembre 2013 12:53
À : Heslop-Harrison, Pat (Prof.); Roland BOURDEIX; Manimekalai; Yaodong Yang; Chandrika Perera; pokou desire; Ramon Rivera; Bee Gunn; harrieshc; Roux, Nicolas (Bioversity-France); Rouard, Mathieu (Bioversity-France); goncalos@uenf.br; thiago.venancio@gmail.com
Objet : Re: Molecular markers in coconut

Le 04/11/2013 21:45, Heslop-Harrison, Pat (Prof.) a écrit :

Some quick thoughts:

"Allele diversity studies as a methodology to assist accession selection" fully agree with this revised section title and rationale.

Later in same section: "cultivar" (and later, population) needs some definition.
Sorry, this is what happens with copy-paste: the terminology changes in the midst of a discussion. Actually, what the software does is independent of the species or of whether we are dealing with cultivars, landraces or natural populations... I opted for "OTUs" (operational taxonomic units.

Then, at what point do you change from representing cultivar diversity to representing allelic diversity? Obviously, not wanting to have representatives of something like a mapping population where each is called a cultivar, but not necessarily discarding some recombinants!

Well, I had not included mapping populations into my reflexion. This needs some thought but spontaneously, I would have treated them separately as "advanced material". Another interesting case would be the synthetic 4-way populations developped in the Philippines.

"attractive or because they represent a standard in genetic trials" good point: do include a good number of weak/susceptible cultivars which are essential to get a handle on the genetic allele effects.
Right, I would not rely on an algorithm that relies only on molecular marker data, even an "optimal" one.

Objective 2: "With the experience acquired. We propose to raise the figure at 10 or 15 per cultivar analysed in Talls and to 5 for Dwarfs." Is this clear evidence that worthwhile extra diversity was found by increasing sample size so much?
The point is not so much to discover extra diversity than to make statistical comparisons easier. This programme (318 individuals in all) is not that large, compared to what would be needed if it was decided to sample each accession in collection (more than some 500 individuals *yearly*)

Two other points: As previously, I would be keen to have preserved soil samples from all sites for future metagenomics analysis - seeing if / where performance/characteristics are related to soil microflora rather than genotype or epigenetics.

Good to think of.

I think there should be some 'future technology' mention: It is sensible to have a focus on what can be done now, but in 3 to 5 years microsats will be superseded. Firstly, when there is a genome sequence, some additional markers may be needed to ensure genome wide coverage, or characterize particular translocations etc. And then later some whole genome (re)sequencing will be more efficient.
An excellent point, thank you.

Best wishes

Pat

Pat Heslop-Harrison.

Professor J.S. (Pat) Heslop-Harrison
Department of Biology
University of Leicester
Leicester LE1 7RH UK

E-mail: phh4@le.ac.uk Skype: Pat.HH Twitter: PatHH1

Annals of Botany blog: www.AoBBlog.com
Websites: www.molcyt.com www.sblab.org
www.bioastral.com www.heslop-harrison.com
(ID/PW: visitor)
Chief Editor, Annals of Botany: www.annbot.com

Phone: +44/0 116 252 5079 / 3381
FAX: +44/0 116 252 2791